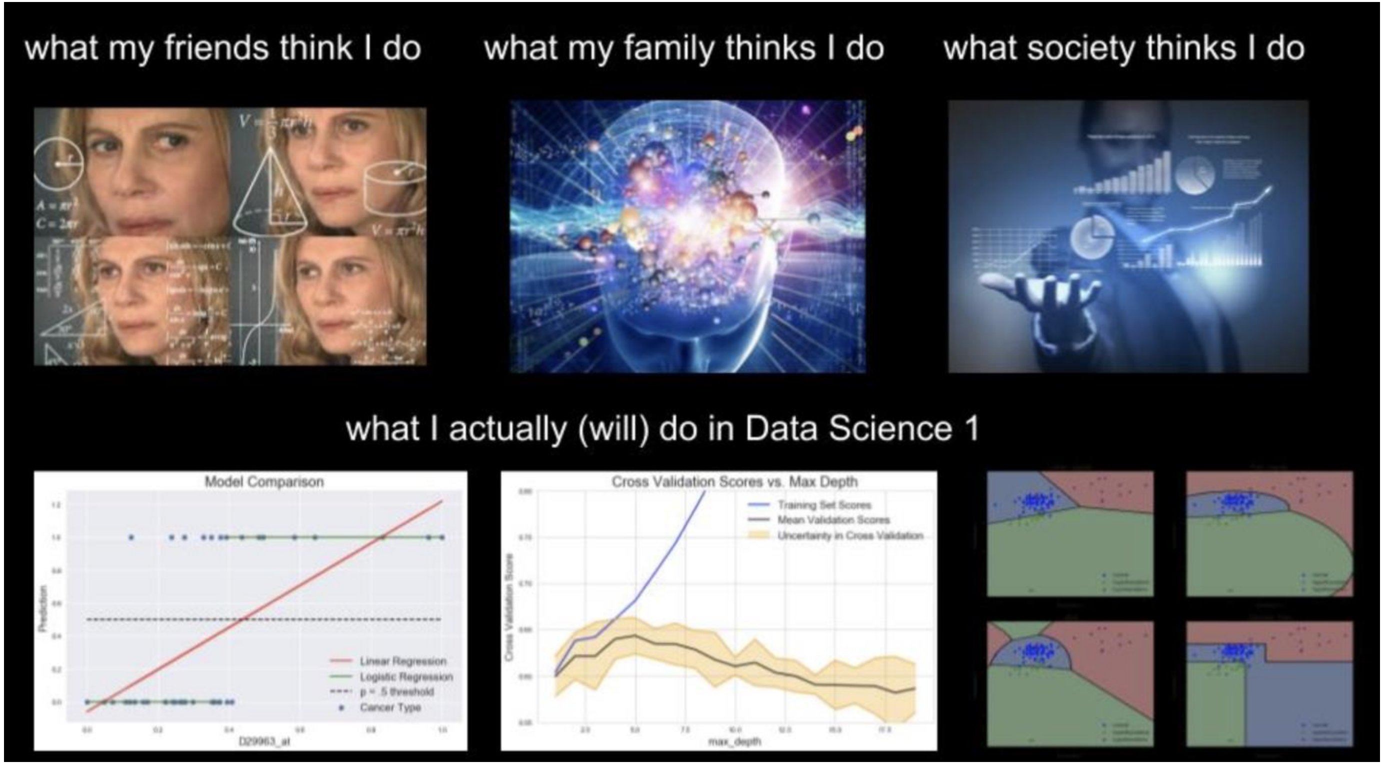
# Introduction Machine Learning

FH University of Applied Sciences TECHNIKUM WIEN

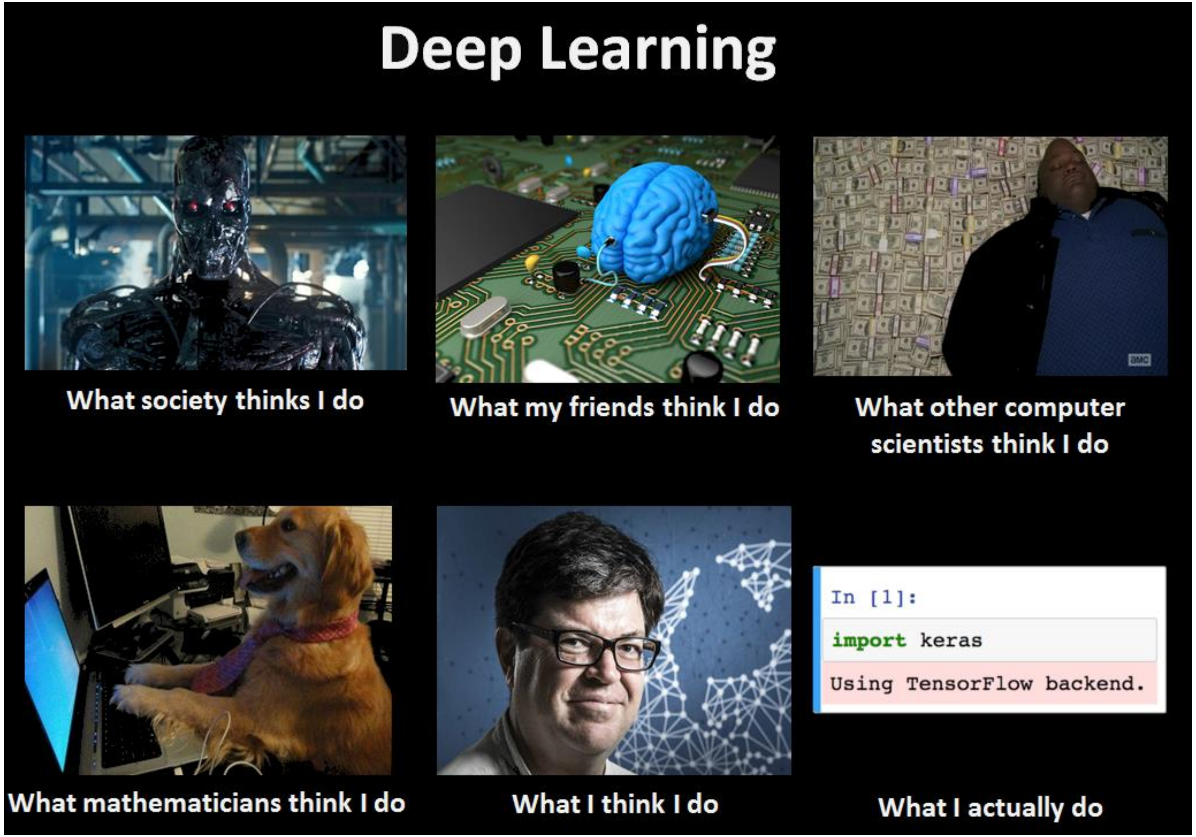
**Lecturer:**

**Authors:** Bernhard Knapp, David Meyer, Pascal Plank, Matthias Blaickner

# Machine Learning? AI? Data Science?



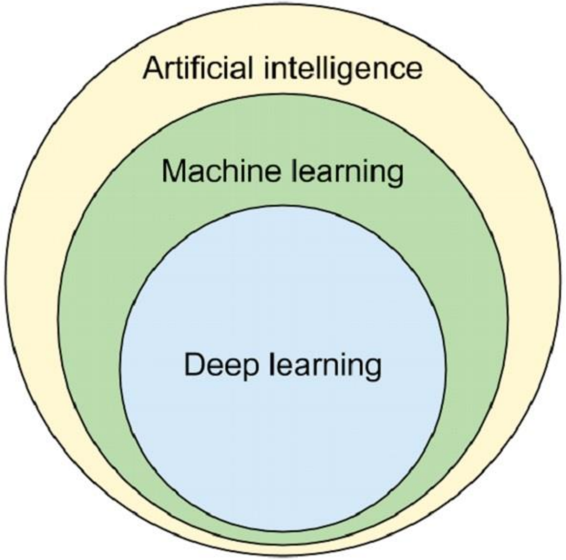
# Here is another meme - sorry I couldn‘t resist!


Deep Learning
What society thinks I do
What my friends think I do
What other computer scientists think I do
What mathematicians think I do
What I think I do
What I actually do

```
In [1]:
import keras
Using TensorFlow backend.
```

# Machine Learning

- “Giving computers the ability to learn from data and to apply that ‘knowledge’ to new data”
- **Aim**: solve a specific or general task optimally without human interference, e. g.
    - classification
    - regression
    - clustering
    - finding abnormalities etc.



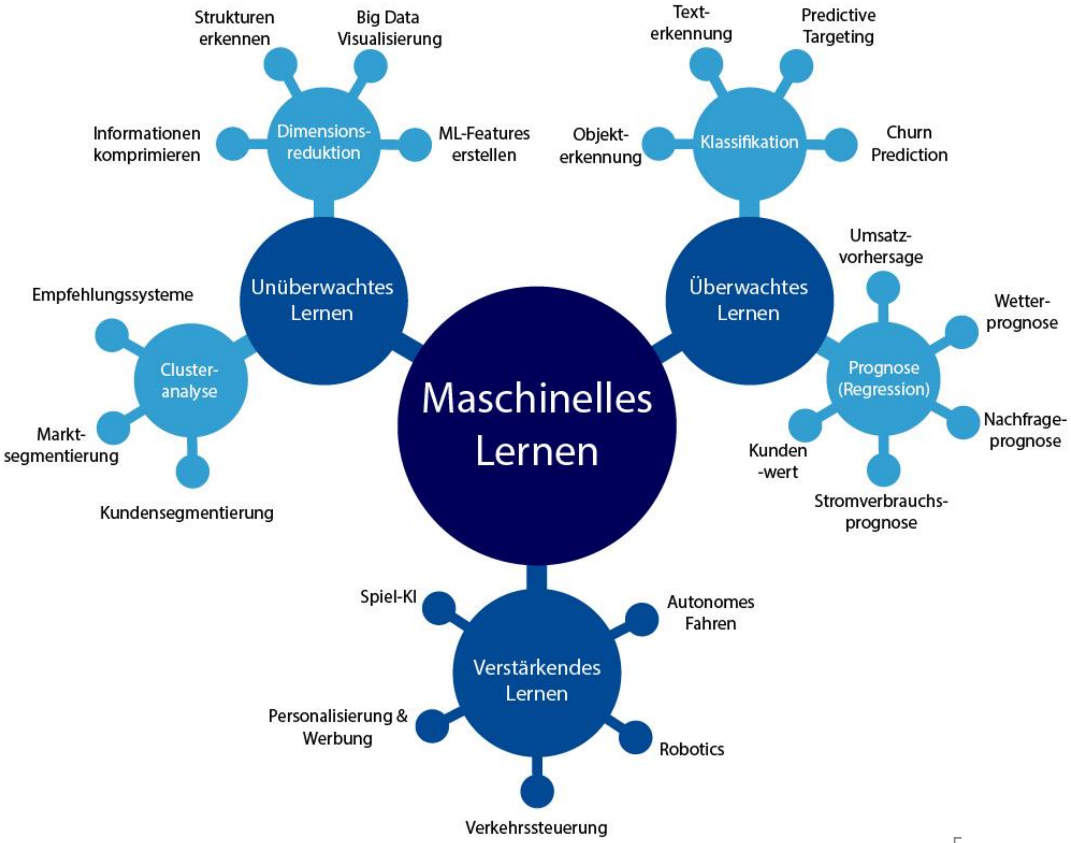
# Machine Learning Types

- **Supervised Learning**:
  - Labelled data
  - Direct feedback

- **Unsupervised Learning**:
  - No labels
  - Finding hidden structures

- **Reinforcement Learning**:
  - Decision process
  - Reward system

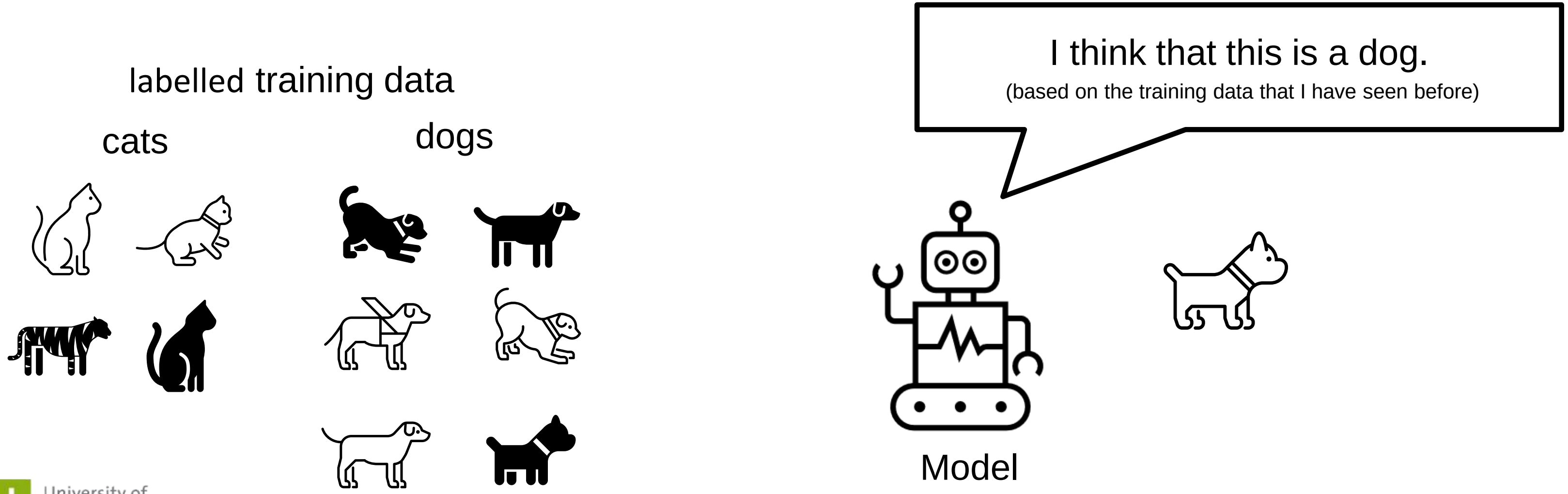
- **Generative AI**
  - Create something new



FH University of Applied Sciences TECHNIKUM WIEN

# Machine Learning Types

- **Supervised Learning**:
    - labelled data
    - Direct feedback
    - Predict an outcome/future, forecasting
    - E. g. predict customers that will return



labelled training data

cats

dogs

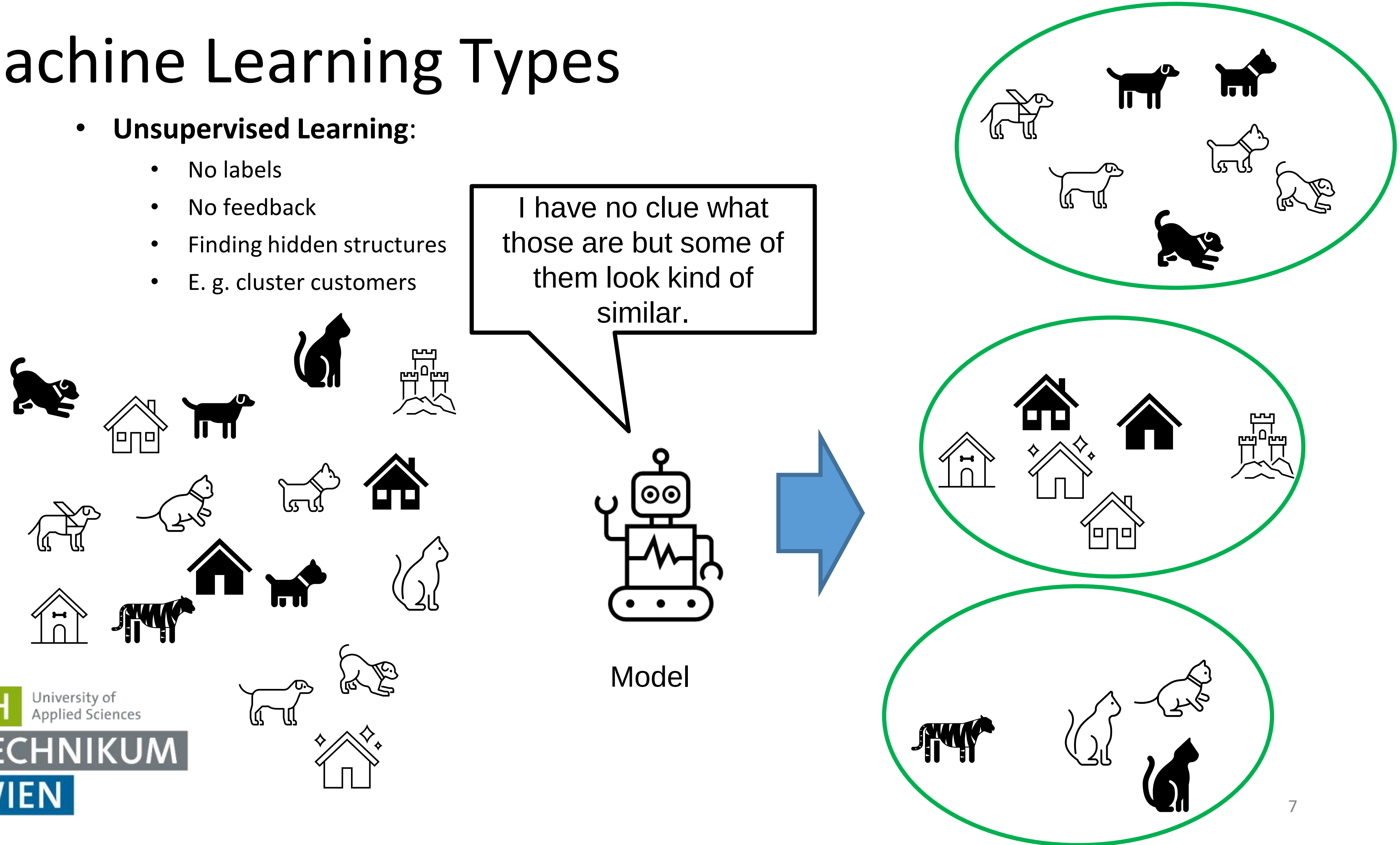
I think that this is a dog.
(based on the training data that I have seen before)

Model

Feedback: correct!

# Machine Learning Types

- **Unsupervised Learning**:
  - No labels
  - No feedback
  - Finding hidden structures
  - E. g. cluster customers

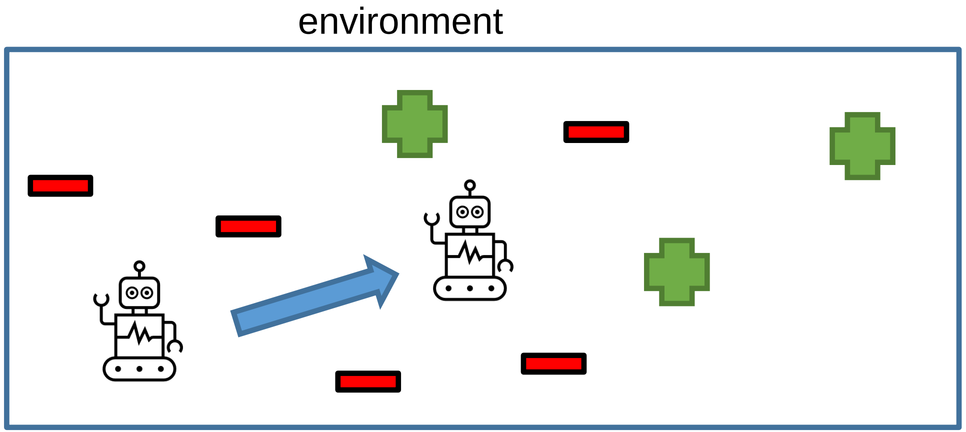


I have no clue what those are but some of them look kind of similar.

Model

# Machine Learning Types

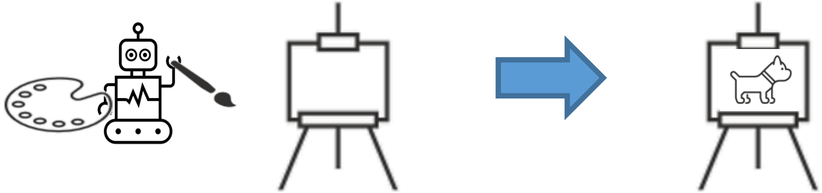
- **Reinforcement Learning**:
  - Decision process
  - Reward system
  - Learn a series of actions
  - E. g. playing Go or Chess

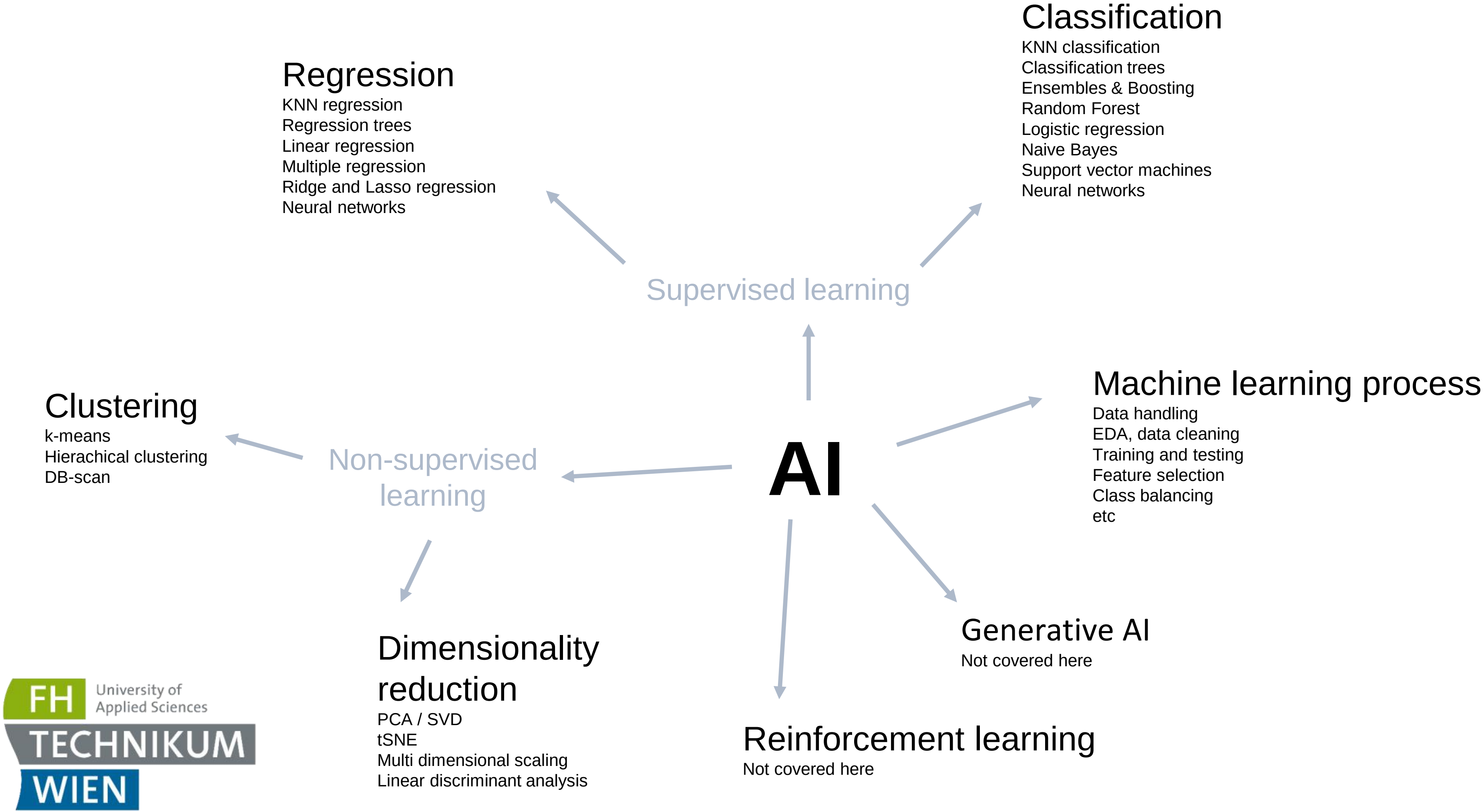


“Agent“ moves around the environment and collects rewards and punishments for its actions

# Machine Learning Types

- **Generative AI**
  - Create something new
  - E.g text, image or song





# AI

Supervised learning

## Regression

KNN regression
Regression trees
Linear regression
Multiple regression
Ridge and Lasso regression
Neural networks

## Classification

KNN classification
Classification trees
Ensembles & Boosting
Random Forest
Logistic regression
Naive Bayes
Support vector machines
Neural networks

Non-supervised learning

## Clustering

k-means
Hierachical clustering
DB-scan

## Dimensionality reduction

PCA / SVD
tSNE
Multi dimensional scaling
Linear discriminant analysis

## Machine learning process

Data handling
EDA, data cleaning
Training and testing
Feature selection
Class balancing
etc

## Generative AI

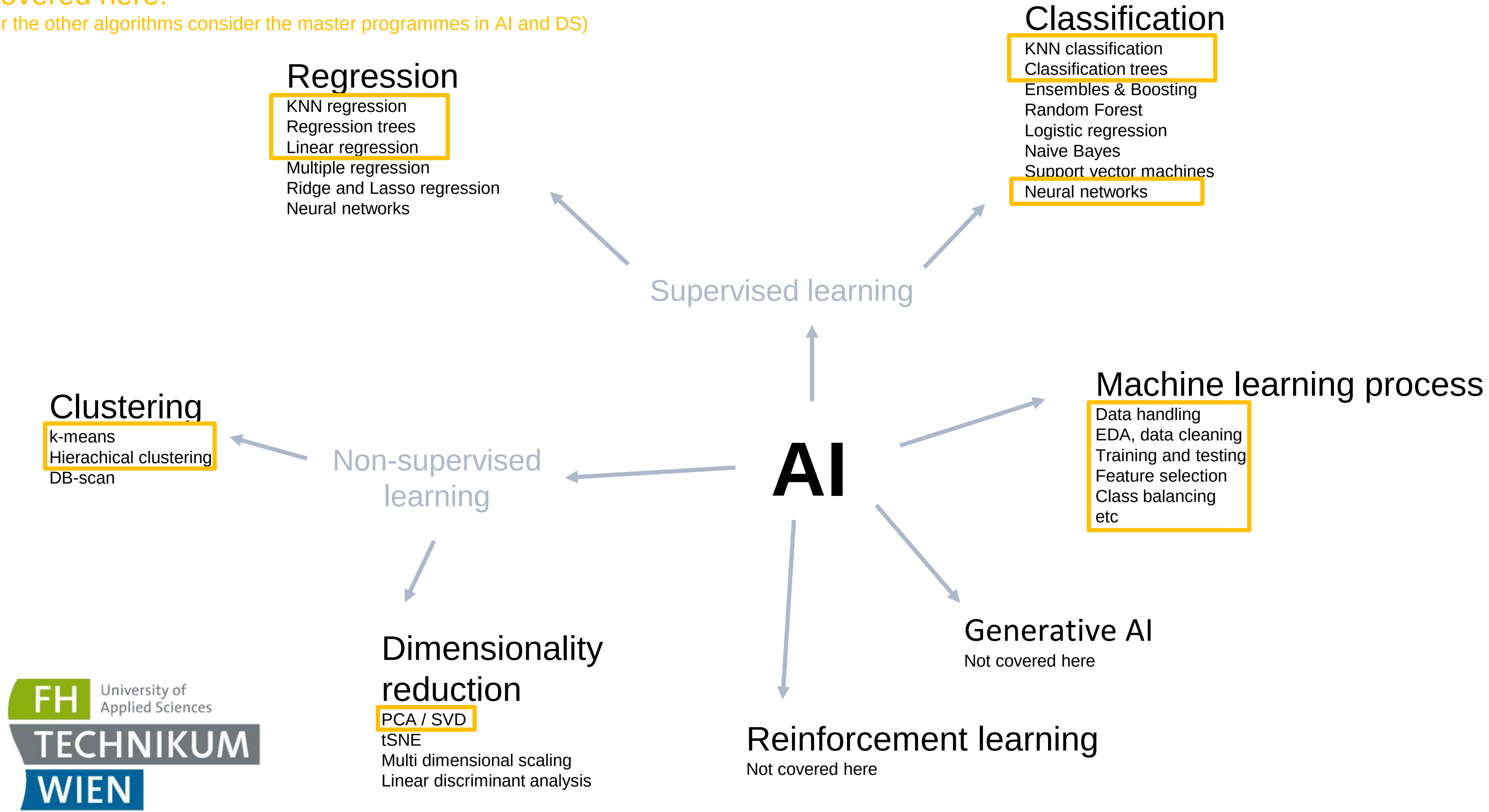
Not covered here

## Reinforcement learning

Not covered here

Covered here:

(for the other algorithms consider the master programmes in AI and DS)



AI

Supervised learning

Regression

KNN regression
Regression trees
Linear regression
Multiple regression
Ridge and Lasso regression
Neural networks

Classification

KNN classification
Classification trees
Ensembles & Boosting
Random Forest
Logistic regression
Naive Bayes
Support vector machines
Neural networks

Non-supervised learning

Clustering

k-means
Hierachical clustering
DB-scan

Dimensionality reduction

PCA / SVD
tSNE
Multi dimensional scaling
Linear discriminant analysis

Machine learning process

Data handling
EDA, data cleaning
Training and testing
Feature selection
Class balancing
etc

Generative AI

Not covered here

Reinforcement learning

Not covered here

# Recent AI Breakthroughs

FH University of Applied Sciences TECHNIKUM WIEN

Name some AI breakthroughs by yourself!

# AI and Games

- 1996: Deep Blue (chess-playing computer developed by IBM) was the first computer to win against a reigning world champion (Garry Kasparov)

- 2016: AlphaGo (Google) AI beats human champion in the much more complex board game “Go” ([https://www.youtube.com/watch?v=WXuK6gekU1Y](https://www.youtube.com/watch?v=WXuK6gekU1Y))

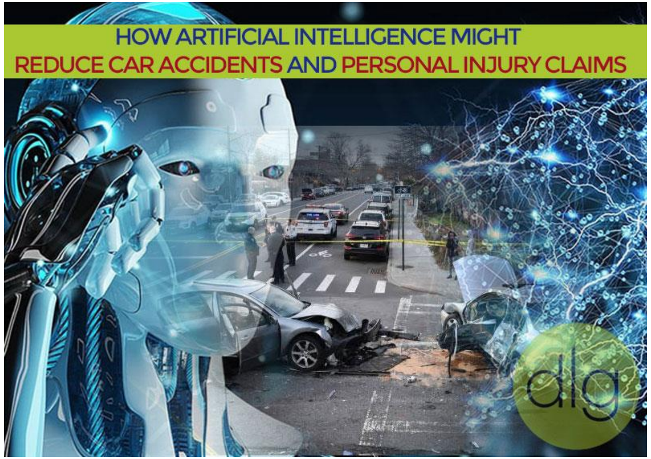
- AI playing computer games:

[https://www.youtube.com/watch?v=cUTMhmVh1qs](https://www.youtube.com/watch?v=cUTMhmVh1qs)

[https://www.youtube.com/watch?v=dJ4rWhpAGFI&t=219s](https://www.youtube.com/watch?v=dJ4rWhpAGFI&t=219s)

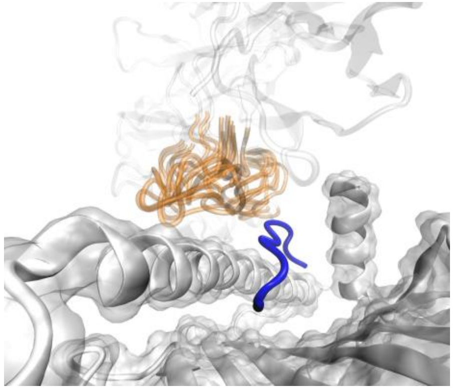
# Self driving cars

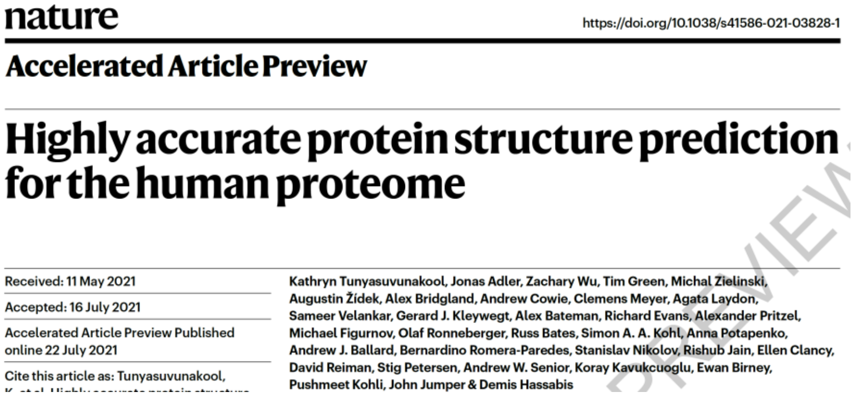
- Try to google for “AI avoids car crash“ e.g. https://www.youtube.com/watch?v=bUhFfunT2ds (start at 45 seconds)


HOW ARTIFICIAL INTELLIGENCE MIGHT
REDUCE CAR ACCIDENTS AND PERSONAL INJURY CLAIMS
dlg

# Prediction of 3D structures of proteins

- Based on known data on how DNA sequences map to protein structures AI learns to produce new protein structures from unfamiliar sequences.
- Google’s AI branch DeepMind launched an algorithm called AlphaFold.
- Google (almost completely) **solved a 50 years old problem**





nature https://doi.org/10.1038/s41586-021-03828-1

Accelerated Article Preview

**Highly accurate protein structure prediction for the human proteome**

Received: 11 May 2021

Accepted: 16 July 2021

Accelerated Article Preview Published online 22 July 2021

Cite this article as: Tunyasuvunakool, K. et al. Highly accurate protein structure

Kathryn Tunyasuvunakool, Jonas Adler, Zachary Wu, Tim Green, Michal Zielinski, Augustin Žídek, Alex Bridgland, Andrew Cowie, Clemens Meyer, Agata Laydon, Sameer Velankar, Gerard J. Kleywegt, Alex Bateman, Richard Evans, Alexander Pritzel, Michael Figurnov, Olaf Ronneberger, Russ Bates, Simon A. A. Kohl, Anna Potapenko, Andrew J. Ballard, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Ellen Clancy, David Reiman, Stig Petersen, Andrew W. Senior, Koray Kavukcuoglu, Ewan Birney, Pushmeet Kohli, John Jumper & Demis Hassabis

# And many more

- Web search engines
- Cleaning robots
- Siri/Alexa
- Diagnostic (medical) AI systems
- Weather forecast
- Smart online shops
- …

# We will not get quite that far …

… but we will learn about **algorithms**, **self implement** algorithms, use **libraries** and hopefully get an understanding of each algorithm as they build the foundation for pretty much every other AI application!

Recommended:

But if you prefer you can use any other type of programming language or library (I am quite agnostic in this aspect)

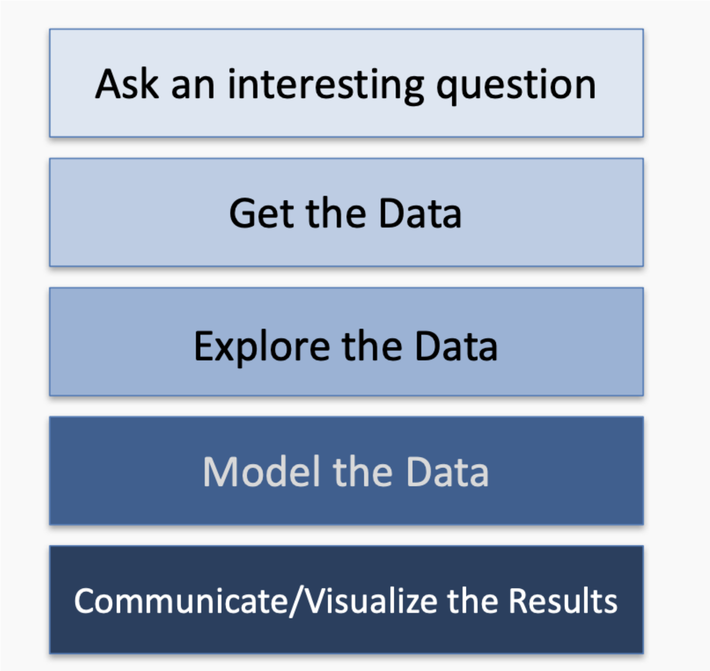
PC PyCharm
The Python IDE

scikit-learn
*Machine Learning in Python*

# The machine learning process

# Data Science

- **The Data Science Process**



Ask an interesting question

Get the Data

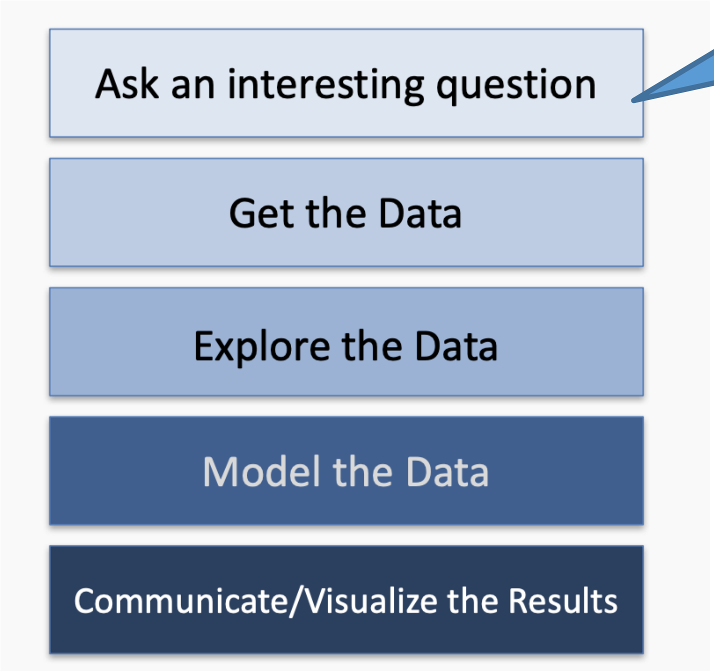
Explore the Data

Model the Data

Communicate/Visualize the Results

# Data Science

- **The Data Science Process**



Ask an interesting question

Get the Data

Explore the Data

Model the Data

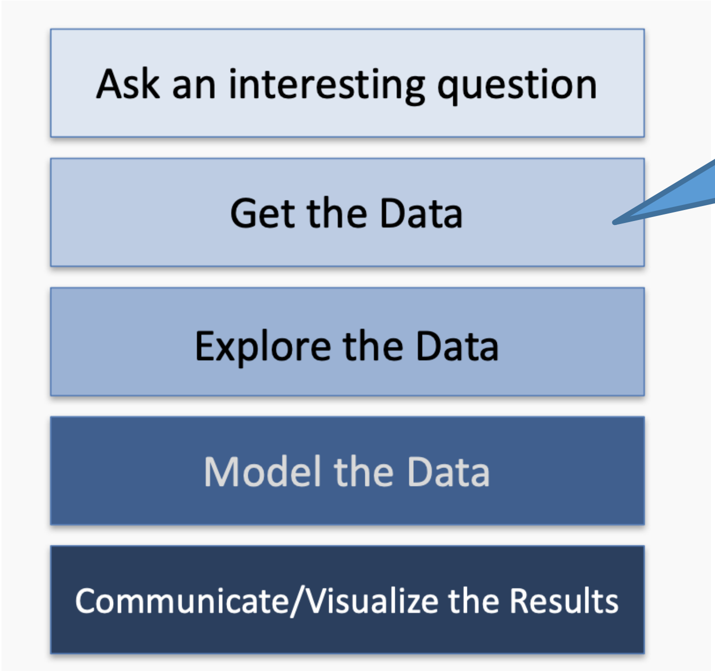
Communicate/Visualize the Results

What is the business goal?
What do you want to predict or estimate?

[process based on Harvard DS course]

# Data Science

- **The Data Science Process**



Ask an interesting question

Get the Data

Explore the Data

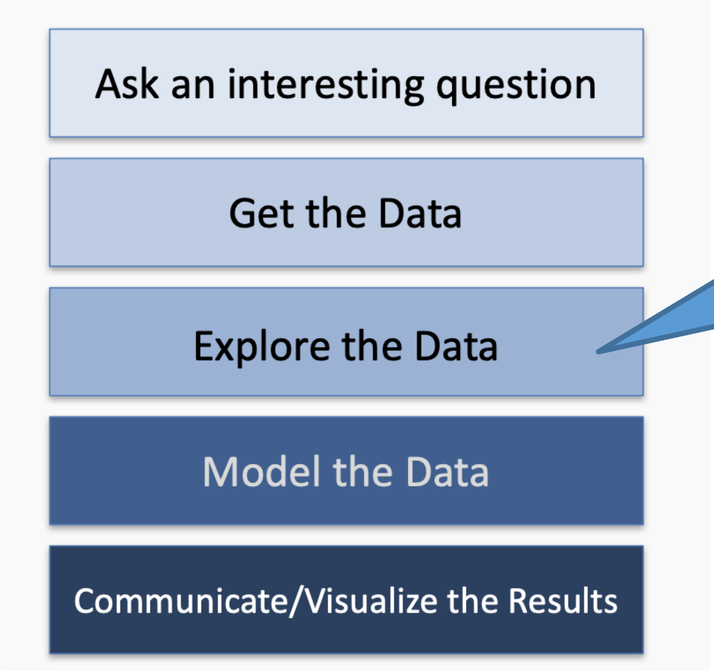
Model the Data

Communicate/Visualize the Results

How were the data sampled?
Which data are relevant?
Privacy / ethical issues?

# Data Science

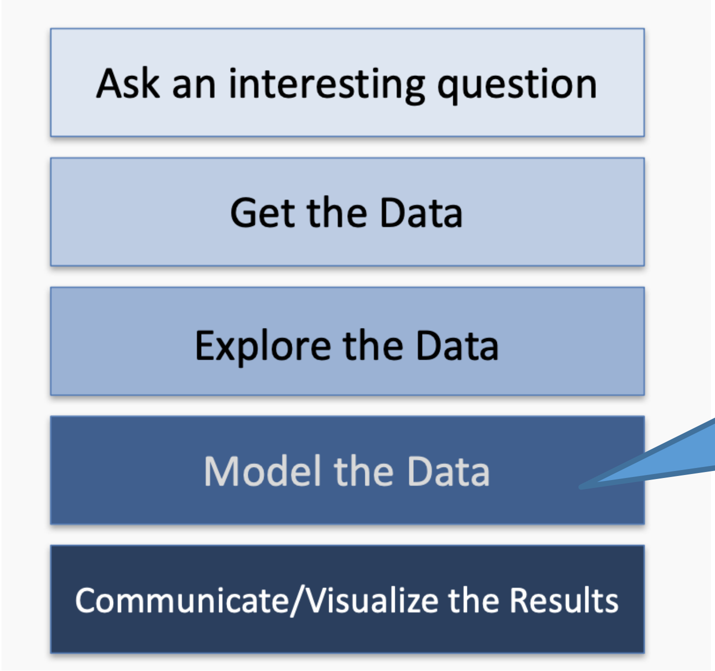
- **The Data Science Process**


Ask an interesting question
Get the Data
Explore the Data
Model the Data
Communicate/Visualize the Results

Get the data.
Plot the data.
Are there anomalies / outliers?
Are there patterns?

# Data Science

- **The Data Science Process**



Ask an interesting question

Get the Data

Explore the Data

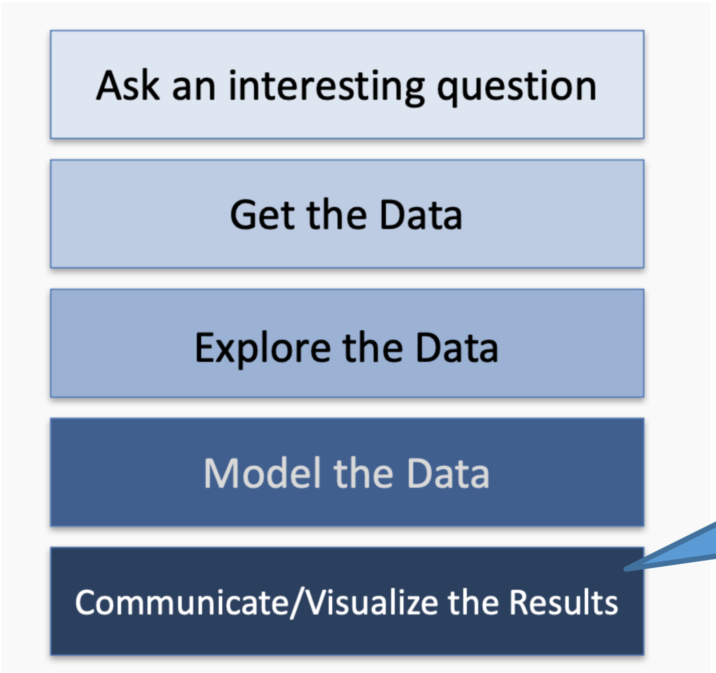
Model the Data

Communicate/Visualize the Results

Build a model.
Fit the model.
Validate the model.
Improve the model? Data?
Try another model?

# Data Science

- **The Data Science Process**



Ask an interesting question

Get the Data

Explore the Data

Model the Data

Communicate/Visualize the Results

What did we learn?
Do the results make sense?
Can we effectively tell a story?

# Course overview

Five key facets of an investigation using data:

1. data collection; data wrangling, cleaning, and sampling to get a suitable data set
2. data management; accessing data quickly and reliably
3. exploratory data analysis; generating hypotheses and building intuition
4. machine learning models
5. communication; summarizing results through visualization, stories, and interpretable summaries.

FH University of Applied Sciences TECHNIKUM WIEN

This is not a linear process!!!

# Course overview

Five key facets of an investigation using data:

1. data collection; data wrangling, cleaning, and sampling to get a suitable data set
2. data management; accessing data quickly and reliably
3. exploratory data analysis; generating hypotheses and building intuition
4. machine learning models
5. communication; summarizing results through visualization, stories, and interpretable summaries.

Covered in this course

# References

[1] AlQuraishi M. End-to-End Differentiable Learning of Protein Structure. Cell Syst. 2019 Apr 24;8(4):292-301.e3. doi: 10.1016/j.cels.2019.03.006.

[2] Philip G Breen, Christopher N Foley, Tjarda Boekholt, Simon Portegies Zwart. Newton versus the machine: solving the chaotic three-body problem using deep neural networks. Monthly Notices of the Royal Astronomical Society, Volume 494, Issue 2, May 2020, Pages 2465–2470. doi.org/10.1093/mnras/staa713

[3] Jonathan A. Weyn, Dale R. Durran, Rich Caruana. Improving Data-Driven Global Weather Prediction Using Deep Convolutional Neural Networks on a Cubed Sphere. Journal of Advances in Modeling Earth Systems 12 (9)2020 doi.org/10.1029/2020MS002109

[4] Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani. An Introduction to Statistical Learning: with Applications in R. New York: Springer, 2013.